Of Mice and Mentors

Developing Cyber-Infrastructure to Support Transdisciplinary Scientific Collaboration

Bradford W. Hesse, PhD

Introduction

When Douglas Engelbart of the Stanford Research Institute (SRI) began refinements on an input device to simplify access to computing systems in 1962, he was setting into motion a cascade of events that would ultimately alter the ways in which scientists worked together. Colloquially, Engelbart referred to his prototype pointing device as a "mouse," a name he gave to the handheld unit when observing that the cord coming out of the back-end looked distinctively similar to a tail (the technical name for the patent was the X-Y Position Indicator for a Display System). Most computer users today recognize the mouse as a mainstay of graphical user computing: a way of pointing, clicking, and dragging "virtual" objects onto either a personal or shared workspace. What users do not recognize is that the invention came out of a radically new way of thinking about knowledge and science.

The Mouse That Roared

What Engelbart and his colleagues set out to do in 1962 was alter the social cognitive environment, or social ecology,[1,2] in which an "augmented"[3] science would take place. Unabashedly, the group had been influenced by the writings of Benjamin Lee Whorf,[4] who suggested that language as a human invention could influence the sophistication of thought: The better and more complete the system for symbolic representation, the better and more sophisticated the intellect it enabled.[4,5] Engelbart and his colleagues reasoned that electronic computer systems represented a natural extension of this thinking, as electronic systems were themselves frameworks for organizing symbolic representations. If the systems could be engineered correctly, they could be used to extend capacity in science. Recognizing that systems and science must co-evolve, the group introduced the term *bootstrapping*[6,7] (literally, to lift oneself up by the bootstraps) to convey a feeling for the iterative course this co-evolution must take.

The mouse was one of the first tools for thought[3] that the group bootstrapped into operation among a select group of scientists in what would come to be known as Silicon Valley.[5] Its purpose was to operate hand-in-hand with a system designed to portray computer data graphically on a screen, and thus give users access and control to a sophisticated set of underlying data patterns in ways that were enlightening and accessible. Using a mouse, the group reasoned, an architect could interact directly with a blueprint for an architectural design on the screen—a metaphor that was more comfortable and understandable than columns of architectural data arrayed in tables.[3] In the context of preventive medicine, an epidemiologist could interact directly with an interactively arrayed map of disease-registry data, looking for disease clusters or signals of outbreak.[8] Both of these ideas may seem commonplace today, but at the time the concept was quite revolutionary.

Another tool introduced by these early cyber-system pioneers was the concept of *hypertext*.[9] The concept was relatively simple. Most language is processed in a linear fashion, but new concepts are formed by making connections between linear strands of logical thought. The hypertext link was introduced as a mechanism for referring a reader to related information instantaneously at the click of the mouse. Although the use of hypertext gained only nominal popularity in personal computing systems, the real power of the mechanism became apparent once the global hypertext linking project, now known as the World Wide Web, matured. Soon, the basic functionality of hypertext was allowing scientists to build off each other's work in unprecedented ways, clicking from one document to the next in pursuit of a hyperlinked thread of continuous thought.

A third defining component of the framework was to enable better collaboration among scientists using online computer-supported cooperative work (CSCW) environments.[10,11] Also called *co-laboratories* (connoting a shared laboratory) or *collaboratories* (connoting a place for online collaboration), these online spaces supported researchers located in different parts of the country and in different time zones as they worked

From the Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland

Address correspondence and reprint requests to: Bradford W. Hesse, PhD, National Cancer Institute, 6130 Executive Boulevard, MSC 7365, Bethesda MD 20892-7365. E-mail: hesseb@mail.nih.gov.

Published by Elsevier Inc. doi:10.1016/j.amepre.2008.05.011

together in virtual space.[12–14] Indeed, completion of the human genome mapping project— one of the most ambitious examples of distributed team science in history—may have been made possible only by the collaborative information infrastructures put in place by biomedical informaticians.

Using Cyber-Infrastructure to Make Team Science Smarter

Early experiments in CSCW environments have had a mixed influence on scientific collaboration.[15,16] On the positive side, scientists who took early advantage of online systems published more prolifically, made more community contacts, and were more successful at requesting use of shared resources than those who were not online.[17] On the negative side, collaborative information environments were clearly not suited for all tasks. Virtual environments could never replace real-world social environments, synchrony, and propinquity in supporting the full gamut of collaborative activities.[13,15,18,19] Regardless of individual costs and benefits, new forms of work began proliferating[16] as individual scientists learned how to query the community as a whole and began coordinating the use of shared, but distant resources in both real and delayed time.[18]

In 2005, authors of a report by the Pew Charitable Trusts declared that online computing—the mouse, hypertext, and computer-supported collaboration—had made its way into the fabric of everyday life.[20] The Internet was no longer an experimental technology waiting for adoption; it was the "new normal." It had insinuated itself as an inseparable dimension of daily work life, and for many professionals it was altering the rules of engagement in substantive and life-altering ways. *New York Times* reporter Thomas Friedman quipped that many of the substantive changes brought about by diffusion of the Internet seemed to be happening "while we were all sleeping"; yet the changes are so monumental they are reshaping the ways in which wealth and power are distributed throughout the world.[21]

Normal science, as a collective enterprise, is experiencing the impact of the new normal firsthand. As Nobel Laureate and Cal Tech President David Baltimore declared when reflecting on changes within the biological community:

> Biology is today an information science. The output of the system, the mechanics of life, are encoded in a digital medium and read out by a series of reading heads. Biology is no longer solely the province of the small laboratory. Contributions come from many directions.[22]

In other words, the fabric of biological science has been permanently altered by the thinking enabled through augmentative information technologies. The life sciences, like many other sciences, are reorganizing themselves along multidisciplinary lines in order to grapple with this new perceived reality.

Grid Computing

One of the core developments in this new era of thinking is the concept of *grid computing*. In April 2005, the American Psychological Association ran a feature article in the *APA Monitor* quoting a University of Chicago professor who observed that the world appears to be quickly dividing into two camps: those who know about grid computing, and those who do not.[23] Those who know about grid computing understand that whole scientific communities have been working to assemble their data structures into an inter-operable lattice of mutually accessible collections of data, tools, and resources.[24] Users of this lattice, or grid, can share resources with each other in order to answer questions that are bigger than what any one single laboratory could solve. Consider how output from thousands of remote sensing devices can be brought together to give geophysicists an unfolding view of global climate change. Or consider how biomedical researchers can channel the terabytes of data collected around the human genome to unlock windows of opportunity for medical intervention. These large-scale, team-science tasks are enabled by the architectures underlying grid computing.[24–26]

Such is the rationale behind the National Cancer Institute (NCI)'s investment in caBIG (the **ca**ncer **B**iomedical **I**nformatics **G**rid).[25,27] Funded originally as an ambitious pilot project, the caBIG infrastructure project is working to provide scientists distributed throughout the NCI's Comprehensive Cancer Centers a common way of accumulating and analyzing data on intracellular processes; clinical manifestations; epidemiologic prevalence, mortality, and incidence; and treatment efficacy. The goal is to accelerate connections in knowledge needed to attack the multi-pronged challenge of cancer from the perspectives of prevention, early detection, diagnosis, treatment, and the long-term management of cancer as a chronic condition.[25,27] Ultimately, the purpose of the caBIG and other grid systems is to co-evolve new tools for thought to match the scope and complexity of science at the beginning of the 21st century. Some of the functionality encompassed by those tools is worth listing.

Transdisciplinary Discovery

New iterations of computer infrastructure, or cyber-infrastructure, are being funded by the National Science Foundation to support the high-performance computing needed to analyze complex, multidisciplinary relationships. The goal is to develop a new evolution of information infrastructure that will be "human-centered, world class, supportive of broadened participation in science and engineering, sustainable, and stable but extensible."[26] Once in place, the expanded resolution of these interconnected and multilevel data sets should open up a new era of discovery in which variables that have never been crossed before are juxtaposed in transdisciplinary analyses.[28] New and advanced data mining techniques are being introduced that can help accelerate the discovery of relationships based on applications of artificial intelligence and machine learning.[29] Understanding the relationship between genes and environment, overcoming health disparities, addressing the multiplex issues of cancer control and prevention are all areas of new discovery enabled by cyber-infrastructure.

Visualization

In the health sciences, efforts are underway to develop tools that can inform the gamut of transdisciplinary analyses from "cells to society."[30] At the cellular level, imaging software is being developed that will allow researchers to visualize macromolecular structures in 3-D, and to manipulate them in real time to reveal hidden aspects of the structure.[26] At the societal level, work is being done by the Open Geospatial Consortium (www.opengeospatial.org/) to develop standards for linking data sets with geographic descriptors. The resulting grid will allow GIS researchers to array anything from disease incidence measures to health knowledge measures geographically on a map.[24] The purpose will be to transform the ways in which health scientists, the public, and policymakers think about complex issues by using the power of cyber-infrastructure to make new graphic relationships accessible through powerful imaging techniques.[31]

Fusion

By some accounts, discussions in the 1970s were focused on the anticipation that there would simply not be enough data to fulfill the promise of advanced computing capabilities. Today, some say, we are "surveying ourselves to death;" that we have more data than we know how to handle and as a result we spend very little of our time integrating findings across data sources.[28] At the very least, this means that we are missing lost opportunities for discovery and decision making. More disconcertedly, we are wasting millions of scarce research dollars on data that are never connected, that never contribute jointly to solving a new but common analytic problem, and that simply stagnate or go unused. Cyber-infrastructure allows for the fusion of related, but heretofore disconnected, data sources.

Decision Support

In previous generations of scientific research, decisions about design and methodology were usually left up to individual researchers operating within isolated laboratories and dependent on the glacial pace of print-text publishing for information from the field. With the advent of the first generation of online collaboratories, scientists began making decisions about the future directions of their research based on the tacit knowledge of scientific colleagues shared online.[12,14] Digital libraries now make it possible to scan the full history of some disciplines with a few simple search terms. Evolution of the digital object identifier (DOI) made it possible for scientists to cross literatures online, jumping through a hyperlink to an online version of an article from the cited reference of another.[32] The development of Web 2.0 technologies (i.e., social computing) is driving this trend further by opening up an online "commons" of scientific knowledge built by volunteers from all stripes and areas of research, the most well known experiment of this type being the online knowledge repository Wikipedia.[33] Similarly, Google Scholar™ is an example of an online search engine that was designed to cross disciplinary silos in retrieving publications.

Policymaking

Changing public policy is often difficult. It requires a honed, persuasive argument relying on credible evidence to persuade and instruct.[34] Once a year, organizers of the Technology, Entertainment, and Design (TED) conference in Monterrey, California invite world-renowned speakers to give "the talk of their lives" (videos are archived and made available to the public at www.TED.com). In February 2006, organizers invited global health expert Hans Rosling to speak at the conference. Using data he had assembled from public health institutions around the world, Rosling gave an engaging presentation that served to shatter audience myths about the nature of poverty, health, and mortality in the Third World. Those data are already driving discussions among policymakers within the European Union, and are generating discussions in policy circles around the globe and illustrate how data synthesis can play an important role in policy change and policymaking. Using the power of connected data sources, scientists can make more compelling arguments to policymakers.

Using Team Science to Make Cyber-Infrastructure More Useful

The promise of grid computing is nothing more than audacious. To create an infrastructure for sharing resources openly in an unfettered information environment across disciplines requires a significant change in culture and incentives. Many less ambitious projects have failed precisely because they did not take into account the incentives and social structures needed to support successful collaboration.[5,11,15] In short, these projects failed, not because of technologic problems, but because network designers failed to heed the lessons learned from team science. In contrast, many examples of success with technologically inferior systems exist precisely because team members were willing to think creatively in devising workarounds for the shortcomings of the technology.[35,36] These projects were successful because of the power of creative collaboration.

The story of cyber-infrastructure, then, lies as much in the study of team science—in collaboration readiness—as it does in the study of new technology—in technology readiness.[13] In this way, the discussions encapsulated in this special issue are especially relevant to the task of building a world-class computer infrastructure for advancing scientific goals. The discussion of evaluation, for example, is directly pertinent to the system designer's ongoing goal of optimizing output. As the science of transdisciplinary evaluation evolves,[37] robust but informative evaluation strategies can be put in place to ensure that the social and technical subsystems[38] of an online science environment work together to meet intended project goals.[15,35]

Likewise, if the benefits from massive data structures interconnected through grid architectures are to materialize, they will come about because of the readiness and willingness of the scientific community to behave in transdisciplinary ways.[37] Research funding agencies and academic policymakers can nurture that process by offering incentives to change the context in which scientific collaboration occurs.[2] Collaborative leaders[39] in preventive medicine can, and should, emerge to help structure the foundations for mass collaboration[33] needed to solve problems of unprecedented complexity in an increasingly connected global environment.

Most importantly, mentors are needed who can take the challenge of modeling new behaviors at a time when the norms of scientific productivity and quality are uncertain. The task will be to move forward with eyes wide open, restructuring their teaching efforts to take full advantage of investments in team science and cyber-infrastructure, while clinging tenaciously to the principles of quality and evidence that must inherently govern scientific collaboration.

No financial disclosures were reported by the author of this paper.

References

1. Montero M, Stokols D. Psychology and the Internet: a social ecological analysis. CyberPsychology & Behavior 2003;6:59–72.
2. Stokols D, Misra S, Moser RP, Hall KL, Taylor BK. The ecology of team science: understanding contextual influences on transdisciplinary collaboration. Am J Prev Med 2008;35(2S):S96–S115.
3. Engelbart DC. Augmenting human intellect: a conceptual framework. Summary Report. Palo Alto CA: Stanford Research Institute, October, 1962.
4. Adams P. Language in thinking: selected readings. Harmondsworth: Penguin, 1972.
5. Bardini T, Friedwald M. Chronicle of the death of a laboratory: Douglas Engelbart and the failure of the knowledge workshop. History of Technology 2002;23:193–212.
6. Engelbart DC. Augment, bootstrap communities, the Web: what next? In: CHI 98 conference summary on human factors in computing systems. Los Angeles, California, United States: ACM Press, 1998.
7. Engelbart DC, Jeff R. Bootstrapping our collective intelligence. ACM Comput Surv 1999;31(4es):38.
8. Ellisman M, Brady M, Hart D, Lin F-P, Miller M, Smarr L. The emerging role of biogrids. Commun ACM 2004;47:52–7.
9. Nelson TH. Complex information processing: a file structure for the complex, the changing and the indeterminate. In: Proceedings of the 1965 20th national conference. Cleveland OH, United States: ACM Press, 1965.
10. Almes G, Cummings J, Birnholtz JP, Foster I, Hey T, Spencer B. CSCW and cyberinfrastructure: opportunities and challenges. In: Proceedings of the 2004 ACM conference on computer supported cooperative work. Chicago IL: ACM Press, 2004.
11. Collins RW. CSCW: an opportunity for action research. In: Proceedings of conference on organizational computing systems. Milpitas, California, United States: ACM Press, 1995.
12. Finholt T. Collaboratories as a new form of scientific organization. Econ Innov New Techn 2003;12:5–25.
13. Olson G, Teasley S, Bietz M, Cogburn D. Collaboratories to support distributed science: the example of international HIV/AIDS research. Proceedings of the SAICSIT 2002:44–51.
14. Finholt T, Olson G. From laboratories to collaboratories: a new organizational form for scientific collaboration. Psychological Science 1997;8:28–35.
15. Sonnenwald DH. Scientific collaboration: a synthesis of challenges and strategies. In: Cronin B, ed. Annual Review of Information Science and Technology. Medford NJ: Information Today, 2007.
16. Sproull L, Kiesler S. Connections: new ways of working in the networked organization. Cambridge MA: MIT Press, 1991.
17. Hesse BW, Sproull L, Kiesler SB, Walsh JP. Returns to science: computer networks in oceanography. Commun ACM 1993;36:90–101.
18. Hesse BW, Werner CM, Altman I. Temporal aspects of computer-mediated communication. Computers in Human Behavior 1988;4:147–65.
19. Olson G, Olson J. Distance matters. Human-Computer Interaction 2000;15:139–78.
20. Pew Foundation. Trends 2005: Information for the Public Interest. Washington DC: Pew Research Center, 2005.
21. Friedman TL. The world is flat: a brief history of the twenty-first century. 1st updated and expanded ed. New York: Farrar, Straus and Giroux, 2006.
22. Denning PJ. Computing is a natural science. Commun. ACM 2007;50:13–8.
23. Greer M. Infrastructure for the science of psychology. Monitor on Psychology 2004:62.
24. Abbas A. Grid computing: a practical guide to technology and applications. Hingham MA: Charles River Media, 2004.
25. Buetow KH. Cyberinfrastructure: empowering a "third way" in biomedical research. Science 2005;308:821–4.
26. National Science Foundation. Cyberinfrastructure Vision for 21st Century Discovery. Arlington VA: National Science Foundation, March 2007.
27. Hesse BW. Harnessing the power of an intelligent health environment in cancer control. Stud Health Technol Inform 2005;118:159–76.
28. Agresti WW. Discovery Informatics. Communications of the ACM 2003;46:25–8.
29. Moore JH. Bioinformatics. J Cell Physiol 2007;213:365–9.
30. Abrams DB. Applying transdisciplinary research strategies to understanding and eliminating health disparities. Health Educ Behav 2006;33:515–31.
31. Bederson B, Shneiderman B. The craft of information visualization: readings and reflections. Amsterdam; Boston: Morgan Kaufmann, 2003.
32. Grudin J. Crossing the divide. ACM Trans Comput-Hum Interact 2004;11:1–25.
33. Tapscott D, Williams AD. Wikinomics: how mass collaboration changes everything. New York: Portfolio, 2006.

34. Nelson DE, Brownson RC, Remington PL, Parvanta C, eds. Communicating public health information effectively: a guide for practitioners. Washington DC: American Public Health Association, 2002.
35. Sonnenwald DH, Whitton MC, Maglaughlin KL. Evaluating a scientific collaboratory: results of a controlled experiment. ACM Trans Comput-Hum Interact 2003;10:150–76.
36. Toffler A, Toffler H. War and anti-war: survival at the dawn of the 21st century. 1st ed. Boston: Little, Brown, 1993.
37. Hall KL, Stokols D, Moser RP, et al. The collaboration readiness of transdisciplinary research teams and centers: findings from the National Cancer Institute TREC year-one evaluation study. Am J Prev Med 2008;35(2S):S161–S172.
38. Pava CH. Managing new office technology: an organizational strategy. New York, NY: Free Press, Collier Macmillan, 1983.
39. Gray B. Enhancing transdisciplinary research through collaborative leadership. Am J Prev Med 2008;35(2S):S124–S132.